Subdifferential Estimates For Marginal Functions Arising From Paramaterized Nonlinear Programming Problems

Supervisor: Dr. Jane Ye
By: Brendan Steed - V00892271

Nov 16, 2020

Abstract

We study the proximal and Frechet subdifferentials of the marginal function derived from a particular class of parameterized nonlinear programming problems. Under an assumption known as inf-compactness, an equivalence between the augmentable Lagrange multipliers of such a problem and the proximal subgradients of the associated marginal function can be shown. For the Frechet subdifferential, we give an upper estiamte in terms of first and second-order Lagrange multipliers.

1 Introduction

For $i \in \{0, 1 \dots, s-1, s, s+1, \dots, m\}$ let $f_i : \mathbb{R}^n \to \mathbb{R}$ and for $u = (u_1, u_2, \dots, u_m)$ consider the parameterized nonlinear programming problem, $P(u)$, given by

$$(P(u)) \qquad \min_{x \in \mathbb{R}^n} \qquad f_0(x)$$

$$\text{s.t.} \qquad f_i(x) + u_i \begin{cases} \leq 0, & i \in \{1, \dots, s\} \\ = 0, & i \in \{s+1, \dots, m\}. \end{cases}$$

We define

$$\mathcal{F}(u) = \left\{ x \in \mathbb{R}^n \,\middle|\, \begin{matrix} f_i(x) + u_i \leq 0, & i \in \{1, \dots, s\} \\ f_i(x) + u_i = 0, & i \in \{s+1, \dots, m\} \end{matrix} \right\} \qquad \text{(Feasible region)}$$

$$p(u) = \inf\{f_0(x) : x \in \mathcal{F}(u)\} \qquad \text{(Marginal function)}$$

$$S(u) = \{x \in \mathcal{F}(u) : f_0(x) = p(u)\} \qquad \text{(Solution set)}$$

under the convention $p(u) = \infty$ when $\mathcal{F}(u) = \emptyset$.

Originally problem $P(u)$ was posited in an economic framework, the parametric dependece of the problem viewed as a resource to be perturbed about some initial $\overline{u} \in \mathbb{R}^m$. However, today, problem $P(u)$ and its marginal function see applications in machine learning, specifically through hyper-parameter selection and bilevel optimization [7].

Denote the first-order Lagrange multipliers for problem $P(u)$ at $x \in \mathcal{F}(u)$ by

$$K^1(u,x) := \left\{ y \in \mathbb{R}^m \;\middle|\; \begin{array}{ll} 0 = \nabla f_0(x) + \sum_{i=1}^{m} y_i \nabla f_i(x) & \\ 0 = y_i(f_i(x) + u_i), & i \in \{1, \ldots, s\} \\ 0 \leq y_i, & i \in \{1, \ldots, s\} \end{array} \right\}.$$

Other than the convex programming case, and some instances in nonconvex programming wherein p happened to be C^2, the first strong results relating the subdifferential theory of p to Lagrange multipliers for solutions to problem $P(u)$ were by Gauvin [5]. Gauvin's analysis was through the somewhat recently defined *Clarke subdifferential* [6] which, in the case of a function f on $\mathbb{R}^m$ which is Lipschitz continuous near $\overline{u} \in \mathbb{R}^m$, is the collection $\partial^C f(\overline{u})$ defined by

$$\partial^C f(\overline{u}) := \{ y \in \mathbb{R}^m : \exists u_n \to \overline{u} \text{ s.t. } \nabla f(u_n) \to y \}.$$

Theorem 1.1 ([5])**.** *Assume p is Lipschitz continuous near $\overline{u} \in \mathbb{R}^m$. Then*

$$\partial^C p(\overline{u}) \subseteq co \left[\bigcup_{x \in S(\overline{u})} K^1(\overline{u}, x) \right].$$

In order to ensure Lipschitz continuity of the marginal function p at some $\overline{u} \in \mathbb{R}^m$, Gauvin assumed that the classical Mangasarian-Fromovitz constraint qualification held at all solution $x \in S(\overline{u})$, as well as assuming a condition with a similar purpose that that of *inf-compactness* (Definition 2.3).

Years later, Rockafellar obtained the same estimate regarding the Clarke subdifferential of p, but without the assumption of MFCQ on problem $P(u)$ [2]. Instead, Rockafellar used a condition analagous to inf-compactness known as *inf-boundedness* [3, Condition (1.1)], together with a growth condition on $P(u)$ (Definition 3.3). However, under his assumptions p was no longer Lipschitz continuous. As such, his estimates are in terms of the full Clarke subdifferential. Eventually, Rockafellar would do away with the growth condition while still maintaining his inf-boundedness assumption [3].

In this paper, we wish to provide estimates for the proximal and Frechet subdifferentials of p in terms of first and second-order Lagrange multipliers for $P(u)$. Section 2 goes over specific notation used in this paper. We also use this section to define the subdifferentials of our analysis, as well as the inf-compactness condition assumed in nearly all results to follow.

The majority of Section 3 is spent proving Theorem 3.1, which is a result of Rockafellar's whose proof spans [1], [2], [3]. Existing estimates for the augmentable Lagrange multipliers

(Proposition 3.2) then yield our desired estimates for the proximal subdifferential in terms of first and second-order Lagrange multipliers.

Section 4 has two results which have not been explicitly seen by the author in the literature. Similar to those given for the proximal subdifferential in Section 3, Lemma 4.1 gives an upper bound for the Frechet subdifferential in terms of first-order Lagrange multipliers while Theorem 4.1 is a bound in terms of the second-order multipliers.

Finally, we use Section 5 to discuss possible weakenings of all results using *restricted inf-compactness* (see Definition 5.1) as opposed to the stronger inf-compactness (see Definition 2.3).

2 Notation and Preliminaries

For $k, q \in \mathbb{N}$, we denote the k-norm of $x = (x_1, x_2, \ldots, x_q) \in \mathbb{R}^q$ by

$$\|x\|_k = \sqrt[k]{\sum_{i=1}^{q} |x_q|^k}.$$

The k-norm δ-ball centered at $x \in \mathbb{R}^q$ will be denoted $B_\delta^k(x) := \{x' \in \mathbb{R}^q : \|x' - x\|_k < \delta\}$. For an operator $A : X \to Y$ where X and Y are Banach spaces, we dentote the operator norm of A as

$$\|A\|_{op} = \sup\{\|Ax\| : x \in X, \|x\| \leq 1\}$$

We will let $C^k(\mathbb{R}^n)$ denote the collection of real-valued functions on $\mathbb{R}^n$ whose k-th partial derivatives all exist and are continuous.

The first subdifferential of importance to this paper is the proximal subdifferential and is defined as follows:

Definition 2.1 (Proximal Subdifferential [4])**.** For $f : \mathbb{R}^n \to [-\infty, \infty]$ a lower-semicontinuous function, the *proximal subdifferential* of f at the point $x \in \mathbb{R}^n$ with $f(x) < \infty$, denoted $\partial^\pi f(x)$, is the collection of vectors $\xi \in \mathbb{R}^n$ such that there exists $\sigma > 0$ and $\delta > 0$ for which

$$f(x') \geq f(x) + \langle \xi, x' - x \rangle - \sigma \|x' - x\|_2^2, \quad \forall x' \in B_\delta^2(x)$$

Observe for $f : \mathbb{R}^n \to \mathbb{R}$ lower-semicontinuous that $\xi \in \partial^\pi f(\overline{x})$ is equivalent to the existence of $g \in C^2(\mathbb{R}^n)$ and $\delta > 0$ satisfying $g(x) \leq f(x)$ for all $x \in B_\delta^2(\overline{x})$, $g(\overline{x}) = f(\overline{x})$, and $\xi = \nabla_x g(\overline{x})$ (See [1]).

One may deduce from the above characterization via C^2 functions that if $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable at $\overline{x}$, then the proximal subdifferential contains only the usual gradient of f at $\overline{x}$. That is, $\partial^p f(\overline{x}) = \{\nabla f(\overline{x})\}$. This is not the case when f is C^1 near $\overline{x} \in \mathbb{R}^n$ (see [4]).

Definition 2.2 (Frechet subdifferental [4]). For $f : \mathbb{R}^n \to [-\infty, \infty]$ lower semicontinuous the *Frechet subdifferential* of f at $x \in \mathbb{R}^n$ with $f(x) < \infty$ is the collection

$$\partial^F f(x) := \left\{ \xi \in \mathbb{R}^n : \liminf_{x' \to x} \frac{f(x') - f(x) - \langle \xi, x' - x \rangle}{\|x' - x\|_2} \geq 0 \right\}.$$

We say $\xi \in \partial^F f(x)$ is a Frechet subgradient.

By definition, $\xi \in \mathbb{R}^n$ is a Frechet subgradient of a lower semicontinuous $f : \mathbb{R}^n \to [-\infty, \infty]$ if and only if for any $\sigma > 0$ there exists a $\delta > 0$ such that

$$f(x') \geq f(x) + \langle \xi, x' - x \rangle + \sigma \|u - \overline{u}\|_2, \quad \forall x' \in B^2_\delta(x).$$

This characterization of the Frechet subgradients will be used extensively to prove our estimate for the Frechet subdifferential in Section 4.

In the context of problem $P(u)$, one has no gaurantee that the marginal function $p : \mathbb{R}^m \to [-\infty, \infty]$ is lower semicontinuous, and hence no gaurentee that there exists well-defined nonempty subdifferential. As such, we employ a condition known as *inf-compactness* to ensure lower semicontinuity of p at particular points of interest.

Definition 2.3 (Inf-compactness [4]). We say that *inf-compactness* holds for problem $P(u)$ at $\overline{u} \in \mathbb{R}^m$ if there exists $\alpha > 0, \delta > 0$ and a bounded set $C \subseteq \mathbb{R}^n$ such that $\alpha > p(\overline{u})$ and

$$\left\{ x \in \mathcal{F}(u) : f_0(x) \leq \alpha, u \in B^2_\delta(\overline{u}) \right\} \subseteq C.$$

Inf-compactness around some $\overline{u} \in \mathbb{R}^m$ also ensures the existence of a solution to $P(\overline{u})$. As stated in the introduction, inf-compactness will be a recurring assumption throughout this paper.

3 Proximal Subdifferential and Augmentable Lagrange Multipliers

We begin this section by defining the (quadratic) augmented Lagrangian for problem $P(u)$.

Definition 3.1 (Augmented Lagrangian). The (quadratic) *augmented Lagrangian* assocaited to problem $P(u)$ is the function $L_u : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}$ defined by

(1) $$L(x, y, r) = f_0(x) + \sum_{i=1}^{s} \varphi\left(f_i(x) + u_i, y_i, r\right) + \sum_{i=s+1}^{m} \psi\left(f_i(x) + u_i, y_i, r\right)$$

where

(2) $$\psi\left(f_i(x) + u_i, y_i, r\right) = y_i[f_i(x) + u_i] + \frac{1}{2} r [f_i(x) + u_i]^2,$$

(3) $$\varphi\left(f_i(x) + u_i, y_i, r\right) = \begin{cases} \psi(f_i(x) + u_i, y_i, r), & \text{if } y_i + r[f_i(x) + u_i] \geq 0 \\ -\frac{1}{2r} y_i^2, & \text{if } y_i + r[f_i(x) + u_i] \leq 0 \end{cases}$$

Note that for any $y \in \mathbb{R}^m$ and $r > 0$, one has [3]

$$L_u(x,y,r) \le f_0(x) \text{ for all } x \in \mathcal{F}(u) \tag{4}$$

and consequently

$$\inf_{x\in C} L_u(x,y,r) \le p(u) \text{ whenever } C \supseteq \mathcal{F}(u) \quad [3]. \tag{5}$$

The associated augmented (or augmentable) Lagrange multipliers can be defined either locally or globally in the following sense.

Definition 3.2. The *globally augmentable Lagrange multipliers* for problem $P(u)$ are given by the collection

$$A_G(u) := \left\{ y \in \mathbb{R}^m : \exists r > 0 \quad \text{such that } \inf_{x\in\mathbb{R}^n} L_u(x,y,r) = p(u) < \infty \right\}. \tag{6}$$

The collection of *(general) augmentable Lagrange multipliers* for problem $P(u)$ is given by

$$A(u) := \left\{ y \in \mathbb{R}^m : \begin{array}{c} \exists r > 0 \text{ and a neighborhood } U \text{ of } u \\ \inf_{x\in F(U)} L_u(x,y,r) = p(u) < \infty \end{array} \right\}. \tag{7}$$

Observe that, as $\mathbb{R}^m$ is open, we have the containment

$$A_G(u) \subseteq A(u) \tag{8}$$

following by definition of a general augmentable Lagrange multiplier.

The main result of this section, whose proof we postpone for the time being, is the following:

Theorem 3.1 ([3])**.** *Assume inf-compactness holds at $\overline{u} \in \mathbb{R}^m$ and that $f_i \in C^0(\mathbb{R}^n)$ for $i \in \{0, 1 \ldots, m\}$. Then*

$$\partial^\pi p(\overline{u}) = A(\overline{u}).$$

First appearing in [2], Theorem 3.1 is a generalization of [1, Theorem 5] wherein $A(u)$ is replaced by $A_G(u)$, and under the assumption of a certain quadratic growth condition for $P(u)$ (Definition 3.3). Note that Theorem 3.1 as stated above is implicitly proven in [3] using results from [1], [2] and [3]. Our goal is to offer a complete proof similar to that which Rockafellar had in mind, but in one location and assuming inf-compactness, as opposed to inf-boundedness.

For the time being, we will require the growth condition used by Rockafellar:

Definition 3.3 (Quadratic Growth Condition [1])**.** The *quadratic growth condition* holds in the context of problem $P(u)$ if

$$\liminf_{\|u\|_2\to\infty} \frac{p(u)}{\|u\|_2^2} > -\infty.$$

A particular case of the quadratic growth condition is when the marginal function p is bounded below over all of $\mathbb{R}^m$. We will see shortly that the quadratic growth condition may be done away with using inf-compactness and considering a slightly modified problem $P(u)$ whose marginal function and augmentable Lagrange multipliers agree with those of the original problem.

Adapted from [1, Theorem 5] is the following lemma, which is the first relation between proximal subgradients and augmented Lagrange multiplier theory for problem $P(u)$ under the quadratic growth condition.

Lemma 3.1. *Assume for $i = 0, 1, \ldots, m$ that $f_i \in C^0(\mathbb{R}^n)$ (i.e. each f_i is continuous). Also assume inf-compactness around $\overline{u}$ and the quadratic growth condition hold for $P(u)$. Then*

$$\partial^\pi p(\overline{u}) = A_G(\overline{u}).$$

Proof. Let $\overline{y} \in A_G(\overline{u})$. Then for some $r > 0$,

$$p(\overline{u}) \leq \inf_{x \in \mathbb{R}^n} L_{\overline{u}}(x, \overline{y}, r). \tag{9}$$

However, by noting [3, Equations $(2.8 - 2.9)$]

$$L_{\overline{u}}(x, y, r) = \min_{u : x \in \mathcal{F}(u)} \left\{ f_0(x) - \langle y, u - \overline{u} \rangle + \frac{1}{2} r \|u - \overline{u}\|_2^2 \right\}$$

we have

$$\inf_{x \in \mathbb{R}^n} L_{\overline{u}}(x, \overline{y}, r) = \inf_{u \in \mathbb{R}^m} \left\{ p(u) - \langle \overline{y}, u - \overline{u} \rangle + \frac{r}{2} \|u - \overline{u}\|_2^2 \right\}. \tag{10}$$

Hence for all $u \in \mathbb{R}^m$,

$$p(u) \geq p(\overline{u}) + \langle \overline{y}, u - \overline{u} \rangle - \frac{r}{2} \|u - \overline{u}\|_2^2$$

so that $\overline{y} \in \partial^\pi p(\overline{u})$.

Now assume $\overline{y} \in \partial^\pi p(\overline{u})$. Then there exists $r_1 > 0$, $\delta > 0$ such that for $u \in B_\delta^2(\overline{u})$,

$$p(u) \geq p(\overline{u}) + \langle \overline{y}, u - \overline{u} \rangle - r_1 \|u - \overline{u}\|_2^2. \tag{11}$$

By definition of the quadratic growth condition, we may find real numbers $r > 0$, and q such that for all $u \in \mathbb{R}^m$

$$p(u) \geq q - r\|u\|_2^2. \tag{12}$$

Choose $r_2 > 0$ so that for all u with $\|u - \overline{u}\| \geq \delta$

$$q - r\|u\|_2^2 \geq p(\overline{u}) + \langle \overline{y}, u - \overline{u} \rangle - r_2 \|u - \overline{u}\|_2^2. \tag{13}$$

Taking $\overline{r} := 2\max\{r_1, r_2\}$ we have from (12) and (13) that for all $u \in \mathbb{R}^m$ with $\|u - \overline{u}\| \geq \delta$

$$p(u) \geq p(\overline{u}) + \langle \overline{y}, u - \overline{u} \rangle - \frac{\overline{r}}{2}\|u - \overline{u}\|_2^2. \tag{14}$$

From (11), we have that for all $u \in \mathbb{R}^m$ with $\|u - \overline{u}\| < \delta$

$$p(u) \geq p(\overline{u}) + \langle \overline{y}, u - \overline{u} \rangle - \frac{\overline{r}}{2}\|u - \overline{u}\|_2^2. \tag{15}$$

Combining (14) and (15), we see

$$p(u) - \langle \overline{y}, u - \overline{u} \rangle + \frac{\overline{r}}{2}\|u - \overline{u}\|_2^2 \geq p(\overline{u}), \quad \forall u \in \mathbb{R}^m$$

Taking infimums in x, we have from (10) that

$$\inf_{x \in \mathbb{R}^n} L_{\overline{u}}(x, \overline{y}, \overline{r}) = \inf_{u \in \mathbb{R}^m} \left\{ p(u) - \langle \overline{y}, u - \overline{u} \rangle + \frac{\overline{r}}{2}\|u - \overline{u}\|_2^2 \right\} \geq p(\overline{u}).$$

Since $\inf_{x \in \mathbb{R}^n} L_{\overline{u}}(x, \overline{y}, \overline{r}) \leq p(\overline{u})$ holds in general, we conclude $\overline{y} \in A_G(\overline{u})$. □

The next proposition is analagous to [3, Proposition 1]. It fullfills a similar role in the proof of Theorem 3.1 to that of the original statement from Rockafellar. In particular, it will serve to gaurentee that when the marginal function p is bounded below globally, then local and global augmentable Lagrange multipliers must agree.

Proposition 3.1. *Assume inf-compactness holds at $\overline{u} \in \mathbb{R}^m$. Let $y \in A(\overline{u})$ and let $B \subseteq \mathbb{R}^m$ be any set such that $\overline{u} \in int(B)$ and p is bounded below on B. Then $y_i \geq 0$ for $i = 1, \ldots, s$ and for all $r > 0$ sufficiently large one has*

$$S(\overline{u}) = \underset{x \in \mathcal{F}(B)}{argmin}\ L_{\overline{u}}(x, y, r) \subseteq int(B). \tag{16}$$

Proof. Recall [3, Equations (2.8 – 2.9)]

$$L_{\overline{u}}(x, y, r) = \min_{u : x \in \mathcal{F}(u)} \left\{ f_0(x) - \langle y, u - \overline{u} \rangle + \frac{r}{2}\|u - \overline{u}\|_2^2 \right\}. \tag{17}$$

Hence, for arbitrary $U \subseteq \mathbb{R}^m$,

$$\inf_{x \in \mathcal{F}(U)} L_{\overline{u}}(x, y, r) = \inf_{u \in U} \left\{ p(u) - \langle y, u - \overline{u} \rangle + \frac{r}{2}\|u - \overline{u}\|_2^2 \right\} \tag{18}$$

with x yielding a minimum on the left if and only if $x \in S(u)$ for some u yielding the minimum on the right (recall that $S(\overline{u}) \neq \emptyset$ by inf-compactness). Taking $U = B$, we see the relation (16) is equivalent to

$$\{\overline{u}\} = \underset{u \in B}{\text{argmin}}\ \left\{ p(u) - \langle y, u - \overline{u} \rangle + \frac{r}{2}\|u - \overline{u}\|_2^2 \right\} \tag{19}$$

since the continuity of the functions $f_i, i \in \{0, 1, \ldots, m\}$ ensures

$$\text{int}\mathcal{F}(B) \supseteq \mathcal{F}(\overline{u}) \supseteq S\overline{u} \text{ when } \overline{u} \in \text{int}(B).$$

On the other hand, the condition $y \in A(\overline{u})$ translates by (18) into the existence of some $r > 0$ and a neighborhood U of $\overline{u}$ such that

$$\overline{u} \in \operatorname*{argmin}_{u \in U} \left\{ p(u) - \langle y, u - \overline{u} \rangle + \frac{r}{2} \|u - \overline{u}\|_2^2 \right\}, \; p(\overline{u}) < \infty. \tag{20}$$

(Note that this implies $y_i \geq 0$ for $i = 1, \ldots, s$ since $p(u)$ is nondecreasing with respect to u_i, $i = 1, \ldots, s$.)

The question therefore boils down to whether (20) holds for some $r > 0$ and neighborhood U ensures that (19) holds for r sufficiently large. Choose $\varepsilon > 0$ such that $\|u - \overline{u}\|_2 \leq \varepsilon$ implies $u \in U$. It suffices to show that if $p(\overline{u})$ is finite and $r_0 > 0$ is such that

$$p(u) \geq p(\overline{u}) + \langle y, u - \overline{u} \rangle - \frac{r_0}{2} \|u - \overline{u}\|_2^2 \text{ when } \|u - \overline{u}\|_2 \leq \varepsilon, \tag{21}$$

then for all $r > 0$ sufficiently large, one will have

$$p(u) > p(\overline{u}) + \langle y, u - \overline{u} \rangle - \frac{r}{2} \|u - \overline{u}\|_2^2 \text{ when } u \in B, \; u \neq \overline{u}. \tag{22}$$

As p is bounded below over B, there is $\alpha \in \mathbb{R}$ such that $p(u) \geq \alpha$ for all $u \in B$. We may also take $\beta \in \mathbb{R}$ such that

$$\langle y, u - \overline{u} \rangle - \frac{r_0}{2} \|u - \overline{u}\|_2^2 \leq \beta \text{ for all } u \in \mathbb{R}^m.$$

If $r > r_0$ but (22) is violated, we would have $\|u - \overline{u}\|_2 > \varepsilon$ but

$$\begin{aligned}
\alpha &\leq p(\overline{u}) + \langle y, u - \overline{u} \rangle - \frac{r}{2} \|u - \overline{u}\|_2^2 \\
&\leq p(\overline{u}) + \langle y, u - \overline{u} \rangle - \frac{r_0}{2} \|u - \overline{u}\|_2^2 - \frac{1}{2}(r - r_0) \|u - \overline{u}\|_2^2 \\
&< p(\overline{u}) + \beta - \frac{\varepsilon}{2}(r - r_0),
\end{aligned}$$

so that $r < r_0 + \frac{2}{\epsilon}[p(\overline{u}) - \alpha + \beta]$. This shows that (22) cannot be violated when r is sufficiently large. □

With Proposition 3.1 in hand, we apply it to discern the when the marginal function is bounded below globally, then local and global augmentable Lagrange multipliers agree.

Lemma 3.2. *If p is bounded below on $\mathbb{R}^m$ and inf-compactness holds at $p(\overline{u})$ then*

$$A(\overline{u}) = A_G(\overline{u}).$$

Proof. Proposition 3.1 assures us that when $y \in A(\overline{u})$, we have for arbitrary $B \subseteq \mathbb{R}^m$ such that p is bounded below on B and $u \in \text{int}(B)$:

$$\inf_{x \in \mathcal{F}(B)} L_{\overline{u}}(x, y, r) = p(u) < \infty \text{ for } r \text{ sufficiently large.} \tag{23}$$

To see this, recall that

$$S(\overline{u}) = \operatorname*{argmin}_{x \in \mathcal{F}(B)} L_{\overline{u}}(x, y, r) \subseteq \text{int } \mathcal{F}(B)$$

if and only if

$$\operatorname*{argmin}_{u \in B} \left\{ p(u) - \langle y, u - \overline{u} \rangle + \frac{r}{2} \|u - \overline{u}\|_2^2 \right\} = \{\overline{u}\}.$$

Then applying equation (18) allows us tot see

$$\inf_{x \in \mathcal{F}(B)} L_{\overline{u}}(x, y, r) = \inf_{u \in B} \left\{ p(u) - \langle y, u - \overline{u} \rangle + \frac{r}{2} \|u - \overline{u}\|_2^2 \right\} = p(\overline{u}).$$

Now, since p is bounded below on $\mathbb{R}^m$ and $\mathcal{F}(\mathbb{R}^m) = \mathbb{R}^n$, we have that in this particular case that Proposition 3.1 implies any $y \in A(\overline{u})$ satisfies

$$\inf_{x \in \mathbb{R}^n} L_{\overline{u}}(x, y, r) = p(\overline{u}) < \infty \text{ for } r \text{ sufficiently large.}$$

Hence,

$$A(\overline{u}) \subseteq A_G(\overline{u}).$$

Recalling that $A_G(\overline{u}) \subseteq A(\overline{u})$ in general to conclude

$$A(\overline{u}) = A_G(\overline{u}).$$

when p is bounded below on $\mathbb{R}^m$. □

Finally, we prove Theorem 3.1. To do so, we construct a modified problem $P(u)$ with the same constraints, but an objective function which is bounded below on $\mathbb{R}^m$, but locally agrees with the objective function of problem $P(u)$ (as a consequence of lower semicontinuity). This local agreement ensures the two problems share the same locally augmentable Lagrange multipliers at $\overline{u}$. Moreover, it also gaurentees that the marginal function of problem $P(u)$ and the marginal function of the modified problem must have the same proximal subdifferential at $\overline{u}$. The theorem then follows by applying the results above to the modified problem.

Theorem 3.2 (Theorem 3.1 Restated). *Assume inf-compactness holds around $\overline{u} \in \mathbb{R}^m$ and that $f_i \in C^0(\mathbb{R}^n)$ for $i \in \{0, 1, \ldots, m\}$. Then*

$$\partial^\pi p(\overline{u}) = A(\overline{u}).$$

Proof. Assume inf-compactness holds around $\overline{u}$ and that $f_i \in C^0(\mathbb{R}^n)$ for $i \in \{0, 1, \ldots, m\}$. Then

$$\partial^\pi p(\overline{u}) = A(\overline{u}).$$

□

Proof. Set $\alpha = p(\overline{u}) - 1$ and consider problem $\tilde{P}(u)$ defined by

$$\left(\tilde{P}(u)\right) \qquad \min_{x\in\mathbb{R}^n} \qquad \tilde{f}_0(x) := \max\{f_0(x), \alpha\}$$
$$\text{s.t.} \qquad f_i(x) + u_i \begin{cases} \leq 0, & i \in \{1, \dots, s\} \\ = 0, & i \in \{s+1, \dots, m\}. \end{cases}$$

as well as the assosciated marignal function $\tilde{p} \geq \alpha$ and augmmented Lagrange multipliers $\tilde{A}(\overline{u})$ and $\tilde{A}_G(\overline{u})$ (defined analagously to p and $A(\overline{u}), A_G(\overline{u})$).

Since $\tilde{p}(u) \geq \alpha$ for all $u \in \mathbb{R}^m$, we have that the quadratic growth condition is satisfied for problem $\tilde{P}(u)$.

Since $\tilde{P}$ also satisfies inf-compactness around $\overline{u}$ and each $f_i \in C^0(\mathbb{R}^n)$, we then have that the functions defining $\tilde{P}(u)$, namely $\tilde{f}_0$ and $f_i, i = 1, \dots, m$, are all continuous. Applying Lemma 3.1 to problem $\tilde{P}(\overline{u})$ we see

$$\partial^\pi \tilde{p}(\overline{u}) = \tilde{A}_G(\overline{u}). \tag{24}$$

As inf-compactness holds around $\overline{u}$ for $P(u)$, the marginal function p is lower semicontinuous at $\overline{u}$. Hence, as $p(\overline{u}) > \alpha$, we have $p(u) > \alpha$ for all u in some neighborhood U_0 of $\overline{u}$. Then for all $x \in \mathcal{F}(U_0)$ we have $f_0(x) > \alpha$ and so $\tilde{f}_0(x) = f_0(x)$. Hence, for $u \in U_0$, we have both

$$\tilde{p}(u) = p(u) \text{ and } \tilde{A}(u) = A(u).$$

Thus, by (24)

$$\partial^\pi p(\overline{u}) = \partial^\pi \tilde{p}(\overline{u}) = \tilde{A}_G(\overline{u}) = \tilde{A}(\overline{u}) = A(\overline{u}).$$

□

Estimates for the proximal subdifferential in terms of more common multipliers for problem $P(u)$ follows by the work of Rockafellar in [3]. It should be noted that there are a plethora of such estimates following from Theorem 3.1, but we give here estimate in terms of the usual first-order multipliers for $P(u)$ and another estimate in terms of somewhat unusual second-order multipliers for problem $P(u)$.

Definition 3.4. Rockafellars second-order Lagrange multipliers for $P(u)$ at $x \in \mathcal{F}(u)$ are given by

$$K^2(u, x) := \left\{ y \in K^1(u, x) \,\middle|\, \begin{array}{l} w \left(\nabla^2_{xx} f_0(x) + \sum_{i=1}^m y_i \nabla^2_{xx} f_i(x) \right) w \geq 0 \\ \forall w \in W(u, x) \end{array} \right\}$$

where $\nabla^2_{xx} f_i(x)$ is the Hessian of f_i at x, and

$$W(u, x) := \left\{ w \in \mathbb{R}^n \,\middle|\, \begin{array}{ll} \langle \nabla f_i(x), w \rangle \leq 0, & \text{for } i \in \{1, \dots, s\} \text{ with } f_i(x) + u_i = 0 \\ \langle \nabla f_i(x), w \rangle = 0 & \text{for } i \in \{s+1, \dots, m\} \\ \quad \nabla f_0(x) \leq 0 & \end{array} \right\}$$

Proposition 3.2 ([3], Proposition 4)**.**

- *Let* $f_i \in C^1(\mathbb{R}^n)$ *for* $i \in \{0, 1 \ldots, m\}$ *and inf-comapctness hold around* $\overline{u}$*. Then for any* $x \in S(\overline{u})$*,*
$$\partial^\pi p(\overline{u}) = A(\overline{u}) \subseteq K^1(\overline{x}, \overline{u}).$$

- *Let* $f_i \in C^2(\mathbb{R}^n)$ *for* $i \in \{0, 1 \ldots, m\}$ *and inf-compactness hold around* $\overline{u}$*. Then for any* $\overline{x} \in S(\overline{u})$*,*
$$\partial^\pi p(\overline{u}) = A(\overline{u}) \subseteq K^2(\overline{x}, \overline{u}).$$

4 Frechet Subdifferential and Lagrange Multipliers

The section discusses estimates for the Frechet subdifferential of the marginal function $p : \mathbb{R}^m \to \mathbb{R}$ for $P(u)$. Lemma 4.1 gives an estimate in terms of usual first-order Lagrange multipliers, $K^1(u, x)$, analagous to the first-order estimate of Proposition 3.2.

Theorem 4.1, while similar to the second-order result of Proposition 3.2 for the proximal subdifferential, we define our Lagrange multipliers in a slightly different manner.

Lemma 4.1. *Assume inf-compactness for problem* $P(u)$ *holds around* $\overline{u} \in \mathbb{R}^m$ *and* $f_i \in C^1(\mathbb{R}^n)$ *for* $i \in \{0, 1, \ldots, m\}$*. Then for each* $\overline{x} \in S(\overline{u})$*,*

$$\partial^F p(\overline{u}) \subseteq K^1(\overline{u}, \overline{x})$$

Proof. Take $\overline{x} \in S(\overline{u})$ (which is nonempty by inf-compactness), $\xi \in \partial^F p(\overline{u})$, and let $\varepsilon > 0$ be given. Then for $\sigma = \min\left\{\epsilon, \frac{\epsilon}{m \sup\{\|\nabla_x f_i(\overline{x})\|_2 : i \in \{1,\ldots,m\}\}}\right\}$, we may find $\delta > 0$ such that

$$p(u) \geq p(\overline{u}) + \langle \xi, u - \overline{u} \rangle - \sigma \|u - \overline{u}\|_2, \quad \forall u \in B^2_\delta(\overline{u}) \tag{25}$$

Rearranging (25), we see

$$p(\overline{u}) \leq p(u) - \langle \xi, u - \overline{u} \rangle + \sigma \|u - \overline{u}\|_2, \quad \forall u \in B^2_\delta(\overline{u}) \tag{26}$$

Note

$$p(u) \leq f(x), \quad \forall x \in \mathcal{F}(u)$$

by definition of $p(u)$. Then since $p(\overline{u}) = f(\overline{x})$ we have from (26) that

$$f(\overline{x}) \leq f(x) - \langle \xi, u - \overline{u} \rangle + \sigma \|u - \overline{u}\|_2, \ \forall u \in B^2_\delta(\overline{u}), \ \forall x \in \mathcal{F}(u).$$

The 2-norm on $\mathbb{R}^m$ is bounded above by the 1-norm so

$$\|u - \overline{u}\|_2 \leq \|u - \overline{u}\|_1 = \sum_{i=1}^m |u_i - \overline{u}_i|.$$

Moreover, by the equivalence of norms on finite dimensional space there exists some $\delta' > 0$ such that $B^1_{\delta'}(\overline{u}) \subseteq B^2_{\delta}(\overline{u})$. Thus for all $u \in B^1_{\delta'}(\overline{u})$ and $x \in \mathcal{F}(u)$ we have

$$f_0(\overline{x}) \leq f_0(x) - \langle \xi, u - \overline{u} \rangle + \sigma \|u - \overline{u}\|_1. = f_0(x) - \sum_{i=1}^{m} \xi_i(u_i - \overline{u}_i) + \sigma \sum_{i=1}^{m} |u_i - \overline{u_i}|.$$

Define a new variable $t = (t_1, t_2, \ldots, t_m) \in \mathbb{R}^m$ and assume for each $i \in \{1, 2, \ldots, m\}$ that $t_i \geq |u_i - \overline{u_i}|$. Then the condition $\sum_{i=1}^{m} t_i \leq \delta'$ ensures $u \in B^1_{\delta'}(\overline{u})$. Hence, for $(x, u, t) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ with $x = (x_1, \ldots, x_m)$, $u = (u_1, \ldots, u_m)$, $t = (t_1, \ldots, t_m)$ and such that

$$\delta' \geq \sum_{i=1}^{m} t_i \tag{27}$$

$$t_i \geq |u_i - \overline{u}_i|, \ \forall i \in \{1, \ldots, m\} \tag{28}$$

$$x \in \mathcal{F}(u) \tag{29}$$

we must have

$$f_0(\overline{x}) \leq f_0(x) - \sum_{i=1}^{m} \xi_i(u_i - \overline{u}_i) + \sigma \sum_{i=1}^{m} t_i. \tag{30}$$

Note that we may replace the m inequalities of (28) by the $2m$ inequalities

$$t_i \geq u_i - \overline{u}_i, \qquad i \in \{1, \ldots, m\} \tag{31}$$

$$t_i \geq -u_i + \overline{u}_i, \qquad i \in \{1, \ldots, m\} \tag{32}$$

Since inequality (30) holds as equality at the point $(x, u, t) = (\overline{x}, \overline{u}, 0)$ (which satisfies conditions (27), (29), and $(31 - 32)$) we have that $(\overline{x}, \overline{u}, 0)$ is a solution to the nonlinear program, $B(\sigma)$ given by

$$(B(\sigma)) \qquad \begin{array}{ll} \min\limits_{(x,u,t)} & f_0(x) - \sum_{i=1}^{m} \xi_i(u_i - \overline{u}_i) + \sigma \sum_{i=1}^{m} t_i \\ \text{s.t.} & -\delta' + \sum_{i=1}^{m} t_i \leq 0 \\ & u_i - \overline{u}_i - t_i \leq 0, \quad \forall i \in \{1, \ldots, m\} \\ & -u_i + \overline{u}_i - t_i \leq 0, \quad \forall i \in \{1, \ldots, m\} \\ & f_i(x) + u_i \leq 0, \quad \forall i \in \{1, \ldots, s\} \\ & f_i(x) + u_i = 0, \quad \forall i \in \{s+1, \ldots, m\} \end{array}$$

The Lagrangian for such a problem is

$$L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^{m-s} \to \mathbb{R}$$

for $(x,u,t,\alpha,\beta,\chi,\eta,\nu) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^{m-s}$ as

$$\begin{aligned} L(x,u,t,\alpha,\beta,\chi,\eta,\nu) =& f_0(x) - \sum_{i=1}^m \xi_i(u_i - \overline{u}_i) + \sigma \sum_{i=1}^m t_i + \alpha\left(-\delta' + \sum_{i=1}^m t_i\right) \\ &+ \sum_{i=1}^m \beta_i(u_i - \overline{u}_i - t_i) + \sum_{i=1}^m \chi_i(-u_i + \overline{u}_i - t_i) \\ &+ \sum_{i=1}^s \eta_i(f_i(x) + u_i) + \sum_{i=1}^{m-s} \nu_i(f_{s+i}(x) + u_{s+i}). \end{aligned}$$

Let us denote

$$\begin{aligned} F_1(x,u,t) &:= -\delta' + \sum_{i=1}^m t_i \leq 0 && \\ F_{(1)_i}(x,u,t) &:= u_i - \overline{u}_i - t_i \leq 0, && \forall i \in \{1,\dots,m\} \\ F_{(2)_i}(x,u,t) &:= -u_i + \overline{u}_i - t_i \leq 0, && \forall i \in \{1,\dots,m\} \\ F_{(3)_i}(x,u,t) &:= f_i(x) + u_i \leq 0, && \forall i \in \{1,\dots,s\} \\ F_{(4)_i}(x,u,t) &:= f_i(x) + u_i = 0, && \forall i \in \{s+1,\dots,m\}. \end{aligned}$$

Under the assumption that f_i is C_1 for each $i \in \{0,1,\dots,m\}$ we have that each of the above functions on $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ are continuously differentiable.

We now show MFCQ holds for $B(\sigma)$ at $(\overline{x},\overline{u},0)$. Note $F_1(\overline{x},\overline{u},0) = -\delta' < 0$ so $F_1(x,u,t)$ is an inactive constraint at $(\overline{x},\overline{u},0)$. Letting e_i denote the i-th basis vector for $\mathbb{R}^{n+m+m}$, we have

$$\begin{aligned} \nabla_{(x,u,t)} F_{(1)_i}(\overline{x},\overline{u},0) &= e_{n+i} - e_{n+m+i} \in \mathbb{R}^{n+m+m} && i \in \{1,\dots,m\} \\ \nabla_{(x,u,t)} F_{(2)_i}(\overline{x},\overline{u},0) &= -e_{n+i} - e_{n+m+i} \in \mathbb{R}^{n+m+m} && i \in \{1,\dots,m\} \\ \nabla_{(x,u,t)} F_{(3)_i}(\overline{x},\overline{u},0) &= \begin{pmatrix} \nabla_x f_i(\overline{x}) \\ 0_{\mathbb{R}^m} \\ 0_{\mathbb{R}^m} \end{pmatrix} + e_{n+i} \in \mathbb{R}^{n+m+m} && i \in \{1,\dots,s\} \\ \nabla_{(x,u,t)} F_{(4)_i}(\overline{x},\overline{u},0) &= \begin{pmatrix} \nabla_x f_i(\overline{x}) \\ 0_{\mathbb{R}^m} \\ 0_{\mathbb{R}^m} \end{pmatrix} + e_{n+i} \in \mathbb{R}^{n+m+m} && i \in \{s+1,\dots,m\}. \end{aligned}$$

Clearly then the collection

$$\left\{\nabla_{(x,u,t)} F_{(3)_i}(\overline{x},\overline{u},0) : i \in \{s+1,\dots,m\}\right\}$$

consisting of gradients (with respect to (x,u,t)) of the equality constraint functions of $B(\sigma)$ evaluated at $(x,u,t) = (\overline{x},\overline{u},0)$ form a linearly independent set of vectors. Moreover, for

$$d = \sum_{i=n+1}^{s} -e_i + \sum_{i=n+m+1}^{n+m+m} 2e_i \in \mathbb{R}^{n+m+m}$$

we have

$$
\begin{aligned}
(e_{n+i} - e_{n+m+i})^T d &< 0, && i \in \{1, \ldots, m\} \\
(-e_{n+i} - e_{n+m+i})^T d &< 0, && i \in \{1, \ldots, m\} \\
\left(\begin{pmatrix} \nabla_x f(\overline{x}) \\ 0_{\mathbb{R}^m} \\ 0_{\mathbb{R}^m} \end{pmatrix} + e_{n+i} \right)^T d &< 0, && i \in \{1, \ldots, s\} \\
\left(\begin{pmatrix} \nabla_x f(\overline{x}) \\ 0_{\mathbb{R}^m} \\ 0_{\mathbb{R}^m} \end{pmatrix} + e_{n+i} \right)^T d &= 0, && i \in \{s+1, \ldots, m\}.
\end{aligned}
$$

Thus MFCQ is satisfied at the solution $(\overline{x}, \overline{u}, 0)$ to $B(\sigma)$. From first order necessary optimality conditions there must exist a first-order Lagrange multiplier $(\alpha, \beta, \chi, \eta, \nu) \in \mathbb{R}_+ \times \mathbb{R}_+^m \times \mathbb{R}_+^m \times \mathbb{R}_+^s \times \mathbb{R}^{m-s}$ satisfying

$$0 = \nabla_{(x,u,t)} L(\overline{x}, \overline{u}, 0, \alpha, \beta, \chi, \eta, \nu) \tag{33}$$

$$0 = \alpha F_1(\overline{x}, \overline{u}, 0) = -\alpha \delta' \tag{34}$$

$$0 = \eta_i F_{(4)_i}(\overline{x}, \overline{u}, 0) = \eta_i \left(f_{s+i}(\overline{x}) + \overline{u}_i \right), \ i \in \{1, \ldots, m-s\} \tag{35}$$

From complementary-slackness, the multiplier corresponding to the constraint $F_1(\overline{x}, \overline{u}, 0) \leq 0$, namely $\alpha \in \mathbb{R}$, is 0. From equation (33) we have

$$\nabla_x L(\overline{x}, \overline{u}, 0, 0, \beta, \chi, \eta, \nu) = 0 = \nabla_x f_0(\overline{x}) + \sum_{i=1}^{s} \eta_i \nabla_x f_i(\overline{x}) + \sum_{i=1}^{m-s} \nu_i \nabla_x f_{s+i}(\overline{x}). \tag{36}$$

$$\nabla_u L(\overline{x}, \overline{u}, 0, 0, \beta, \chi, \eta, \nu) = 0 = -\xi + \beta - \chi + \begin{pmatrix} \eta \\ \nu \end{pmatrix} \tag{37}$$

$$\nabla_t L(\overline{x}, \overline{u}, 0, 0, \beta, \chi, \eta, \nu) = 0 = \sigma \mathbf{1}_{\mathbb{R}^{m \times 1}} - \beta - \chi \tag{38}$$

where $1_{\mathbb{R}^{m \times 1}} = \sum_{i=1}^{m} e_i \in \mathbb{R}^m$. In particular, $\eta_i = \xi_i + \chi_i - \beta_i$ for $i \in \{1, \ldots, s\}$ and $\nu_{i-s} = \xi_i + \chi_i - \beta_i$ for $i \in \{s+1, \ldots, m\}$. This combined with equation (36) shows

$$0 = \nabla_x f_0(\overline{x}) + \sum_{i=1}^{m} \xi_i \nabla_x f_i(\overline{x}) + \sum_{i=1}^{m} (\chi_i - \beta_i) \nabla_x f_i(\overline{x}).$$

However, since $\sigma = \beta_i + \chi_i$ and $\beta_i, \chi_i \geq 0$ for all $i \in \{1, \ldots, m\}$ it holds then that

$$|\beta_i - \chi_i| \leq |\beta_i| + |\chi_i| = \beta_i + \chi_i = \sigma < \frac{\epsilon}{m \sup \{\|\nabla_x f_i(\overline{x})\|_2 : i \in \{1, \ldots, m\}\}}.$$

Hence,

$$\left\|\nabla_x f_0(\overline{x}) + \sum_{i=1}^m \xi_i \nabla_x f_i(\overline{x})\right\|_2 \leq \left\|\sum_{i=1}^m (\chi_i - \beta_i)\nabla_x f_i(\overline{x})\right\|_2$$
$$\leq \sum_{i=1}^m \|(\chi_i - \beta_i)\nabla_x f_i(\overline{x})\|_2$$
$$\leq m\sigma \sup\left\{\|\nabla_x f_i(\overline{x})\|_2 : i \in \{1,\ldots,m\}\right\} < \epsilon.$$

As this holds for every $\varepsilon > 0$, we conclude $\nabla_x f_0(\overline{x}) + \sum_{i=1}^m \xi_i \nabla_x f(\overline{x}) = 0$.

Also for every $i \in \{1,\ldots,s\}$ we have $\eta_i \geq 0$ so from $0 \leq \chi_i, \beta_i \leq \sigma$

$$\xi_i = \eta_i + \beta_i - \chi_i \geq -\sigma > -\epsilon.$$

Since this holds for every $\epsilon > 0$ we conclude $\xi_i \geq 0$ for all $i \in \{1,\ldots,s\}$.

Finally, $f_i(\overline{x}) + \overline{u}_i < 0$ for some $i \in \{1,\ldots,s\}$ implies $\eta_i = 0$ by complementary slackness. Therefore, in this case, we have another estimate,

$$|\xi_i| = |\xi_i - \eta_i| = |\beta_i - \chi_i| \leq \sigma \leq \epsilon.$$

Again, as this holds for all $\epsilon > 0$, we conclude $\xi_i = 0$ for all $i \in \{1,\ldots,s\}$ such that $f_i(\overline{x}) + \overline{u}_i < 0$.

As

$$\begin{aligned} 0 &= \nabla_x f_0(\overline{x}) + \sum_{i=1}^m \xi_i \nabla_x f_i(\overline{x}) \\ 0 &= \xi_i \left(f_i(\overline{x}) + \overline{u}_i\right), && i \in \{1,\ldots,s\} \\ \xi_i &\geq 0, && i \in \{1,\ldots,s\} \end{aligned}$$

we conclude $\xi \in K^1(\overline{u}, \overline{x})$. □

On the other hand, Theorem 4.1 gives an estimate for the Frechet subdifferential in terms of the usual second-order multipliers for problem $P(u)$ given here:

Definition 4.1. The usual second-order multipliers for $P(u)$ at $x \in \mathcal{F}(u)$ are given by

$$M^2(u,x) := \left\{ y \in K^1(u,x) \,\middle|\, \begin{array}{l} d^T \left(\nabla^2_{xx} f_0(x) + \sum_{i=1}^m y_i \nabla^2_{xx} f_i(x)\right) d, \\ \forall d \in \Omega(u,x) \end{array} \right\},$$

where

$$\Omega(u,x) := \left\{ d \in \mathbb{R}^n \,\middle|\, \nabla f(x)^T d = 0, \forall i \in \{1,\ldots,m\} \text{ with } f_i(x) + u_i = 0 \right\}.$$

Theorem 4.1. *Assume inf-compactness for $P(u)$ holds around $\overline{u} \in \mathbb{R}^m$, and assume $f_i \in C^2(\mathbb{R}^n)$ for $i \in \{0, 1, \ldots, m\}$. Then for every $\overline{x} \in S(\overline{u})$,*

$$\partial^F p(\overline{u}) \subseteq M^2(\overline{u}, \overline{x}).$$

Proof. Let $\xi \in \partial^F p(\overline{u})$. Take $\overline{x} \in S(\overline{u})$. Then as $f_i \in C^2(\mathbb{R}^n) \subset C^1(\mathbb{R}^n)$ we have $\xi \in K^1(\overline{u}, \overline{x})$.

Assume to the contrary that there exists some $d \in \Omega(\overline{u}, \overline{x})$ such that

$$d^T \left(\nabla^2_{xx} f_0(\overline{x}) + \sum_{i=1}^{m} \xi_i \nabla^2_{xx} f_i(\overline{x}) \right) d = a < 0.$$

Then for $\sigma \in \mathbb{R}$ with

$$0 < \sigma < \frac{-a}{m \cdot \|d\|_2^2 \cdot \sup_{i \in \{1,\ldots,m\}} \|\nabla^2_{xx} f(\overline{x})\|_{op}}$$

we have by the reformulation in the proof of Lemma 4.1 that $(\overline{x}, \overline{u}, 0)$ is a solution to $B(\sigma)$ for which MFCQ holds. By the second order necessary conditions, there exists a second-order multiplier (of the usual kind),

$$(\alpha, \beta, \chi, \eta, \nu) \in \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^{m-s}$$

corresponding to the solution $(\overline{x}, \overline{u}, 0)$ of $B(\sigma)$. In particular, $(\alpha, \beta, \chi, \eta, \nu)$ is a first-order Lagrange multiplier for the solution $(\overline{x}, \overline{u}, 0)$ of $B(\sigma)$ which also satisfies

$$v^T \nabla^2_{(x,u,t)} L_0(\overline{x}, \overline{u}, 0, \alpha, \beta, \chi, \eta, \nu) v \geq 0$$

for all $v \in \mathbb{R}^{n+m+m}$ such that

$$\left(\nabla_{(x,u,t)} F_{(j)_i}(\overline{x}, \overline{u}, 0)\right)^T v = 0, \qquad j \in \{1, 2\},\ i \in \{1, \ldots, m\} \tag{39}$$

$$\left(\nabla_{(x,u,t)} F_{(3)_i}(\overline{x}, \overline{u}, 0)\right)^T v = 0, \qquad i \in \{1, \ldots, s\} \text{ with } F_{(3)_i}(\overline{x}, \overline{u}, 0) = 0 \tag{40}$$

$$\left(\nabla_{(x,u,t)} F_{(4)_i}(\overline{x}, \overline{u}, 0)\right)^T v = 0, \qquad i \in \{s+1, \ldots, m\}, \tag{41}$$

where $\nabla^2_{(x,u,t)} L : \mathbb{R}^{n+m+m} \to \mathbb{R}^{n+m+m}$ is the Hessian of the Lagrangian L of $B(\sigma)$ with respect to the first $n+m+m$ coordinates corresponding to (x, u, t).

Notice

$$\begin{aligned}
&\nabla^2_{(x,u,t)} L(\overline{x}, \overline{u}, 0, \alpha, \beta, \chi, \eta, \nu) \\
&= \begin{pmatrix}
\nabla^2_{xx} f_0(\overline{x}) + \sum_{i=1}^{s} \eta_i \nabla^2_{xx} f_i(\overline{x}) + \sum_{i=1}^{m-s} \nu_i \nabla^2_{xx} f_{s+i}(\overline{x}) & | & 0_{n\times m} & | & 0_{n\times m} \\
- & & - & & - \\
0_{m\times n} & | & 0_{m\times m} & | & 0_{m\times m} \\
- & & - & & - \\
0_{m\times n} & | & 0_{m\times m} & | & 0_{m\times m}
\end{pmatrix}
\end{aligned}$$

and that for any $d \in \Omega(\overline{u}, \overline{x})$ we have

$$v = \begin{pmatrix} d \\ 0_{m\times 1} \\ 0_{m\times 1} \end{pmatrix} \in \mathbb{R}^{n+m+m}$$

satisfies equations $(39-41)$ simultaneously. As $(\alpha, \beta, \chi, \eta, \nu)$ is a second-order Lagrange multiplier for the solution $(\overline{x}, \overline{u}, 0)$ of $B(\sigma)$, we then have

$$(42) \qquad \begin{pmatrix} d \\ 0_{m\times 1} \\ 0_{m\times 1} \end{pmatrix}^T \nabla_{xx} L_0(\overline{x}, \overline{u}, 0, \alpha, \beta, \chi, \eta, \nu) \begin{pmatrix} d \\ 0_{m\times 1} \\ 0_{m\times 1} \end{pmatrix} = d^T \left(\nabla^2_{xx} f_0(\overline{x}) + \sum_{i=1}^{s} \eta_i \nabla^2_{xx} f_i(\overline{x}) + \sum_{i=1}^{m-s} \nu_i \nabla^2_{xx} f_{s+i}(\overline{x}) \right) d \geq 0$$

As $(\alpha, \beta, \chi, \eta, \nu)$ is a first-order Lagrange multiplier for the solution $(\overline{x}, \overline{u}, 0)$ of $B(\sigma)$, we have $\alpha = 0$ and that equations $(36-38)$ still hold. These equalities combined with equation (42) allows us to see

$$\begin{aligned}
0 &\leq d^T \left(\nabla^2_{xx} f_0(\overline{x}) + \sum_{i=1}^{m} \xi_i \nabla^2_{xx} f_i(\overline{x}) \right) d + d^T \left(\sum_{i=1}^{m} (\chi_i - \beta_i) \nabla^2_{xx} f_i(\overline{x}) \right) d \\
&= a + d^T \left(\sum_{i=1}^{m} (\chi_i - \beta_i) \nabla^2_{xx} f_i(\overline{x}) \right) d \\
&\leq a + \left| d^T \left(\sum_{i=1}^{m} (\chi_i - \beta_i) \nabla^2_{xx} f_i(\overline{x}) \right) d \right| \\
&\leq a + \sum_{i=1}^{m} (\chi_i - \beta_i) \left\| \nabla^2_{xx} f_i(\overline{x}) \right\|_{op} \cdot \|d\|_2^2 \\
&= a + \sigma \|d\|_2^2 \sum_{i=1}^{m} \left\| \nabla^2_{xx} f_i(\overline{x}) \right\|_{op} \\
&\leq a + \sigma \cdot m \cdot \|d\|_2^2 \cdot \sup_{i \in \{1,\ldots,m\}} \left\| \nabla^2_{xx} f_i(\overline{x}) \right\|_{op} \\
&< 0
\end{aligned}$$

by our choice of $\sigma > 0$. Hence $0 < 0$, a contradiction.

We conclude

$$d^T \left(\nabla^2_{xx} f_0(\overline{x}) + \sum_{i=1}^{m} \xi_i \nabla^2_{xx} f_i(\overline{x}) \right) d \geq 0$$

for all $d \in \Omega(\overline{u}, \overline{x})$.

□

5 Concluding Remarks

As mentioned in the introduction, a condition strictly weaker than inf-compactness, known as *restricted inf-compactness* [8, Definition 3.8] is defined by:

Definition 5.1. We say that the *restricted inf-compactness* holds around $\overline{u}$ if $p(\overline{u})$ is finite and there exists a compact set C and a positive number ε_0 such that, for all $u \in B^2_{\varepsilon_0}(\overline{u})$ for which $p(u) < p(\overline{u}) + \varepsilon$, the problem $P(u)$ has a solution in C.

The author is fairly convinced that all of the analysis above is valied assuming restircted inf-compactness. The key reason being that inf-compactness is only applied to ensure lower semicontinuity of the value function, as well as nonemptiness of the solution set. However, if using the Frechet or proximal subdifferentails to construct the limiting subdifferential, there is a chance that restricted inf-compactness becomes too weak and that even a weaker upper bound of the form $\partial^L p(\overline{u}) \subseteq \bigcup_{x\in S(\overline{u})}$, where $\partial^L p(\overline{u})$ is the limiting subdifferential of p at $\overline{u}$. The main concern of the author is that restricted inf-compactness at $\overline{u}$ does not imply the nonemptiness of $\mathcal{S}(\overline{u})$ near $\overline{u}$.

Another interesting consideration of the author would be the possibility of a proof for the proximal subdifferential estimates in terms of Lagrange multipliers which uses a method analagous to the Frechet subdifferential estimates proof of Lemma 4.1 and Theorem 4.1. By introducing a new viarable, there is an opportunity to define a smooth differentiable programming problem using inequality from the proximal subgradient definition.

Finally, I would like to thank Dr. Ye and Dr. MacGillivray for their invaluable help in writing this thesis.

References

[1] R. T. Rockafellar, Augmented Lagrange Multiplier Functions and Duality in Nonconvex Programming, *SIAM Journal on Control*, **12(2)** (1974), 286–285

[2] R. T. Rockafellar, Proximal Subgradients, Marginal Values, and Augmented Lagrangians in Nonconvex Optimization, *Mathematics of Operations Research* **6(3)** (1981), 424–436.

[3] R.T. Rockafeller, Marginal Values and Second-Order Necessary Conditions for Optimality, *Mathematical Programming* **25** (1983), 245–286.

[4] R. T. Rockafellar and R. J. Wets, Variational Analysis, *Springer Verlag* (1998) Heidelberg, Berlin, New York.

[5] J. Gauvin, The Generalized Gradient of a Marginal Value Function in Mathematical Programming, *Mathematics of Operations Research* **4** (1979), 458–463.

[6] F. H. Clarke, Generalized Gradients and Applications, *Trans. Amer. Math. Soc.* **205** (1975), 247–262.

[7] J. J. Ye, On Solving Bilevel Optimization Problems, *Presented at a workshop on "Variational Analysis and Optimization"*, UBC Okanagan, Kelowna (2019).

[8] L. Guo, G. Lin, J.J. Ye, Z. Zhang, Sensitivity Analysis of the Value Function For Parametric Mathematical Programs with Equilibrium Constraints, *SIAM J. Optimal Control* **24** (2014), 1206–1237